

CLAIMS

What is claimed is:

1 1. An apparatus for filtering documents as received, the apparatus  
2 comprising:  
3 a document parser, said document parser accepting documents as input and  
4 providing inverted lists of terms contained in the documents as output;  
5 a profile parser, the profile parser accepting user queries as input as  
6 providing a query net representing the query as output; and  
7 a comparator that compares the inverted list for an incoming document  
8 against the query net representing the user query and providing as output an  
9 indication whether the incoming document matches the user query.

1 2. The apparatus of claim 1 wherein said profile parser provides as output a  
2 term dictionary containing terms from the user query, and wherein the document  
3 parser uses said term dictionary to eliminate terms from said inverted list.

1 3. The apparatus of claim 1 wherein said profile parser accepts a plurality of  
2 user queries and stores a corresponding plurality of query nets in a memory  
3 element.

1 4. The apparatus of claim 3 wherein said comparator compares an inverted  
2 list associated with an incoming document against each of said plurality of stored  
3 query nets.

1 5. The apparatus of claim 1 wherein said comparator provides an indication  
2 that the incoming document matches the user query if said inverted lists contain a  
3 minimum number of terms contained in the query net.

1 6. The apparatus of claim 1 wherein said profile parser, said document parser,  
2 and said comparator reside on separate machines, said separate machines  
3 interconnected by a network.

Sub A2  
1 A method for filtering incoming documents, the method comprising the  
2 steps of:

3 (a) receiving an incoming document and parsing it to produce an  
4 inverted list of terms contained in incoming document;

5 (b) using the produced inverted list to retrieve query nets representing  
6 user queries;

7 (c) discarding retrieved query nets matching less than a predetermined  
8 number of terms;

9 (d) scoring remaining profile and discarding profiles having a score less  
10 than a predetermined threshold.

1 8. The method of claim 7 further comprising the step of receiving user queries  
2 and parsing the user queries to produce query nets representing the query.

1 9. The method of claim 7 further comprising that step of storing all query nets  
2 associated with a user as a clipset.

1 10. The method of claim 7 further comprising the step of storing an inverted  
2 list associated with an incoming document.

1 11. The method of claim 9 further comprising the step of storing corpus  
2 statistics in the clipset.

Sub A3  
1 12. An article of manufacture having computer-readable program means for  
2 filtering incoming documents, the article comprising:

- 3 (a) computer-readable program means for receiving an incoming  
4 document and parsing it to produce an inverted list of terms contained in incoming  
5 document;  
6 (b) computer-readable program means for using the produced inverted  
7 list to retrieve query nets representing user queries;  
8 (c) computer-readable program means for discarding retrieved query  
9 nets matching less than a predetermined number of terms;  
10 (d) computer-readable program means for scoring remaining profile  
11 and discarding profiles having a score less than a predetermined threshold.

1 13. The article of manufacture of claim 12 further comprising computer-  
2 readable program means for receiving user queries and parsing the user queries to  
3 produce query nets representing the query.

1 12. 14. The article of manufacture claim <sup>11</sup>12 further comprising computer-readable  
2 means of storing all query nets associated with a user as a clipset.

1 14. 15. The article of manufacture of claim <sup>11</sup>12 further comprising  
2 computer-readable means of storing an inverted list associated with an incoming  
3 document.

a 1 13. 16. The article of manufacture of claim <sup>12</sup><sub>14</sub> further comprising computer-readable  
2 means of storing corpus statistics in the clipset.

Add B4

0891257.084897  
458T80" 5/27/88